**1)** **For the "Crowdsourcing Annotations for Visual Object Detection" paper, identify and describe each of the crowdsourcing microtasks (in 1-3 sentences per microtask) that are used in the complete crowdsourcing system pipeline to detect objects. [9 points]**

**Ans.** The sub tasks used in the crowdsourcing pipeline to detect objects are:

### 1. Drawing Task

- In this task a crowd worker is asked to draw a bounding box around one instance of a given image. It is important to ensure that the crowd worker works on images which have at least one object instance that has not yet been covered with bounding box.

- The workers who work on the drawing task need to undergo a worker training and pass a qualifying test. The crowd worker needs to read a set of instructions (rules) in order to draw accurate and precise bounding box .The paper mentions 4 different rules that need to be followed performing the drawing tasks.

  Rule 1: The bounding box should be as tight as possible and should include all visible part of object.

  Rule 2: If a single image has multiple instances of an object, then only one instance should be included in one bounding box.

  Rule 3: If an object instance already has a bounding box around it, then worker should draw the next bounding box on new instance.

  Rule 4: If the image doesn't have the required object instance or the instance has already been marked, then the worker needs to check a 'checkbox' on screen stating the same.

- The worker needs to read the instructions and give a test in order to qualify for working on the drawing task. They are given instant feedbacks to improve their annotating speed and accuracy. The quality for this task is controlled by the quality verification task.

### 2. Quality Verification Task

- In quality verification task the worker is asked to verify the quality of bounding boxes that were made in the previous step. To maintain good quality control, they just have a single bounding box per image that needs to be verified.

- Similar to drawing task, this task also requires the worker to undergo training and pass a qualification test to start working on real quality verification tasks.

- In the training, the workers are asked to read and understand a set of rules to get an understanding of "good bounding box".

  Rule 1: The bounding box includes the instance of required object.

  Rule 2: It should be as tight as possible and must include all visible parts.

  Rule 3: If we have an image with multiple object instances, then good bounding box must include just one instance of the desired object.

- The quality of these verification tasks is controlled by embedding "gold standard" ie each task has a set of validation images that contains good and bad images. The good validation bounding boxes are collected by majority voting.

3. **Coverage Verification Task**

- In the Coverage task the crowd worker verifies if all the instances of an object is rounded by a bounding box.

- Similar to drawing and quality verification task, the workers are trained and need to pass a qualification test so that they can work on this task.

- The quality control for this task is similar to that of "quality verification task", two type of validation images are created; one that has bounding box around all the occurrences of object instance and the other in which the all the instances are not covered. The quality is ensured by evaluating the performance of worker in correctly identifying validation images.

  (Su, 2012, p.3) (Su, 2012, p.4) (Su, 2012, p.5)

2) **For the "Large Scale Image Annotations on Amazon Mechanical Turk" paper, describe each the crowdsourcing system for "Keypoint Annotation of Objects".[6 points]**

**Ans:** In Keypoint Annotation of Objects, the crowd sourcing task involves marking the various joints (referred as keypoints) that are present in an image. The author of the paper mentions two challenges that need to be addressed for successfully accomplishing the crowd sourcing task.

- The first challenge is deciding on the reference keypoint that will be used for images from varied categories. Since all the objects have different orientation and structures, it is difficult to define accurate keypoint for every possible object. For eg. as mentioned in paper it is easier to define keypoints for object belonging to animal categories as the anatomical part can be used as keypoints, but for objects from categories like plane have varied structural differences. (Maji, 2011, p.3)

- In order to get a solution for the first challenge, the authors decided to split the categories into common sub-categories. Each sub-category now had a set separate keypoints for its object.

- The second challenge was that not all the categories had principal orientation, due to which it was difficult to assign keypoints in the reference frame of given object. To handle this, the approach was to have view dependent keypoints.

**Annotation task by crowd workers**

- The crowd workers have a user interface that has the image with a bounding box in it. They are given set of instructions and a list of key points that needs to be referred while marking the key points in the image. The interface also has list of examples to help the crowd worker accurately mark the keypoints.

- If a worker is not able to locate a keypoint, they can just leave it at its initial position on the interface. It is also possible to reset the annotation and start from beginning if the worker feels he made any mistake in the process. It is important that the user/worker marks only visible points while annotating. (Maji, 2011, p.6)

**Quality Verification**

- To ensure good quality control each object is annotated by five different crowd worker. A keypoint is considered to be present in an object if at least two workers correctly mark it in the image. The location of keypoint is determined by finding the closest pair of annotations and averaging entire list annotation which are present within certain radius of them.

**3) Choose one of the crowdsourcing systems you discussed from questions 1 and 2, and describe two ways that you would improve the crowdsourcing system. [5 points]**

**Ans**: The two ways in which I believe we can improve the crowdsourcing system - "Keypoint Annotation of Object" is:

- Firstly, I think instead of using the bounding box in the annotation task, we can draw a closed polygon around the instance of the object that needs to be annotated. For eg. If we have images with multiple instances of an object, and the some of the instances are overlapping to some extent, then in this case having a polygonal coverage around the instance will be more efficient and help us restrict our boundary to our desired instance.

  Consider fig 4. in the paper, which has two horses in it, standing close to each other. In this case if we have a bounding box then it may happen that it covers at least some part of the other horse. This will raise an ambiguity when the crowd workers mark the keypoints. Having a closed polygonal boundary will help us to restrict our boundary only to single instance while dealing with overlapping or closely placed object instance images.

- Secondly, I believe that the verification task can be made better. Suppose we have an image with higher level of complexity like overlapping or closely placed object instances, in such tasks it may happen that crowd worker is unable to mark the keypoints accurately. And many crowd workers face this difficulty, resulting in bad annotation at a high level. In this case if we are just check if two annotators have marked a keypoint, then it becomes difficult to ensure that if it is actually present or not.

  To improve this, we can use a verification step similar to that in "Figure Ground masks of objects" ie having pair wise overlap between the keypoints present in an object and finding the one with overlaps with maximum with others. We can also define a threshold value for objects with greater complexity. Also, instead of just 5 independent users per image, we can increase it to probably 7-8 user to get more accurate annotations.

**4) In addition, please submit one "discussion point" in total (not per paper). This can be in the form of a question, critique, connection to other readings, or plausible future work that you think is interesting to investigate in greater detail in class. The discussion point may be about one paper or can compare and contrast different papers. In addition, the discussion point can address the proposed idea, methods, experimental design, and analysis of results. This should be roughly 1 sentence. long [5 points]**

**Ans:** I believe that in the "Keypoint Annotation of Object", if we have a complex image with multiple instances of an object, it may be difficult for crowd workers to correctly annotate the key points which may result in many crowd workers incorrectly marking the keypoints. In such cases, I feel that instead of just having 5 independent annotation per image we can increase this number and can have 10 workers working on same complex image. I want to discuss if it is an efficient way to improve the quality control?